

AD \_\_\_\_\_

Award Number: DAMD17-94-J-4076

TITLE: Development of a Common Database for Digital Mammography Research

PRINCIPAL INVESTIGATOR: Robert M. Nishikawa, Ph.D.

CONTRACTING ORGANIZATION: The University of Chicago  
Chicago, Illinois 60637

REPORT DATE: October 2001

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;  
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20021115 034

**REPORT DOCUMENTATION PAGE**Form Approved  
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

**1. AGENCY USE ONLY (Leave blank)****2. REPORT DATE**

October 2001

**3. REPORT TYPE AND DATES COVERED**

Final (15 Sep 94 - 14 Sep 01)

**4. TITLE AND SUBTITLE**

Development of a Common Database for Digital Mammography Research

**5. FUNDING NUMBERS**

DAMD17-94-J-4076

**6. AUTHOR(S)**

Robert M. Nishikawa, Ph.D.

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**The University of Chicago  
Chicago, Illinois 60637

E-Mail: r-nishikawa@uchicago.edu

**8. PERFORMING ORGANIZATION  
REPORT NUMBER****9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012**10. SPONSORING / MONITORING  
AGENCY REPORT NUMBER****11. SUPPLEMENTARY NOTES****12a. DISTRIBUTION / AVAILABILITY STATEMENT**

Approved for Public Release; Distribution Unlimited

**12b. DISTRIBUTION CODE****13. ABSTRACT (Maximum 200 Words)**

The purpose of this infrastructure project is to develop a large database of digitized mammograms that will be distributed free of charge to researchers working in all aspects of digital mammography. This database will facilitate and promote rapid development in digital mammography research. The database will consist of 1000 cases subdivided into 5 categories, 4 containing different breast lesions -- masses, microcalcifications, architectural distortions, asymmetric densities (both benign and malignant) -- and one containing normal mammograms. The mammograms will be collected and digitized (0.05-mm pixel size) at two sites: the Universities of Chicago and North Carolina. The database will be stored at the two sites and will be available over internet, and by mail on CD, tape, and magneto-optical disks. To date 630 cases have been digitized. Each case consists of index and previous exams (each having four standard views) and up to two special-view mammograms (e.g., magnification views). The computer systems for the database have been assembled and are connected to the network. The first release of 50 cases with clustered microcalcifications will be made shortly after true location of the cluster is determined. This release will be followed by another 50 cases of microcalcifications and 100 cases with masses. Along with the images, instructions on how to use the database will also be distributed. This includes an objective scoring method that we are developing based on input from radiologists'.

**14. SUBJECT TERMS**

Digital Mammography, Database, Information Systems, Image Analysis, Computer-aided Diagnosis, Image Processing

**15. NUMBER OF PAGES**

8

**16. PRICE CODE****17. SECURITY CLASSIFICATION  
OF REPORT**

Unclassified

**18. SECURITY CLASSIFICATION  
OF THIS PAGE**

Unclassified

**19. SECURITY CLASSIFICATION  
OF ABSTRACT**

Unclassified

**20. LIMITATION OF ABSTRACT**

Unlimited

### **3. TABLE OF CONTENTS**

Front Cover .....	1
SF 298 Report Documentation .....	2
Table of Contents .....	3
Introduction .....	4
Body .....	4
Key Research Accomplishments .....	6
Reportable Outcomes .....	6
Conclusions .....	6
References .....	6

#### **4. INTRODUCTION**

The purpose of this project is to develop a database of digital mammograms that can be used by researchers who (1) are trying to determine the image quality requirements of detectors for digital mammography; (2) are developing image processing techniques to optimize the displayed digital mammogram; (3) are developing computerized methods for analyzing mammograms; (4) are studying the effects of image compression methods on image quality; (5) are developing methods for remote transmission of mammograms; and (6) are studying the relationship between image quality and diagnostic accuracy. This database also could be used as a resource for teaching radiology residents and for testing the performance levels of mammographers.

The specific aims of this proposal are:

1. Collect and digitize 200 cases in each of 5 different categories, mammograms exhibiting: (i) clustered microcalcifications, (ii) masses, (iii) architectural distortions, (iv) asymmetric densities, and (v) no lesions (i.e. normals).
2. Make these cases available to other researchers either over computer network (Internet) or by sending images on computer tape or CD. The database will be distributed as widely as possible so that comparisons of different computerized analysis techniques can be standardized.

#### **5. BODY**

This research is being funded as an infrastructure award and as such, it does not represent a research project per se. That is, there is no hypothesis that we are trying to prove. Therefore, this report is structured slightly different from a normal scientific research report -- heavy on the method and light on actual results. In this project, the procedure is the most important component, which is applied continuously in a straightforward manner to achieve the goal of creating the database of mammograms.

##### **Task 1: Collect and digitize mammograms, (See Figure 1.)**

We now have 630 cases digitized (see Table I, at the end of the report). Most of these cases still need to be truth marked accurately. Currently, because of the high volume of clinical work in the radiology department, it has been difficult to get a radiologist to mark the truth. Two research-oriented radiologists (Drs. Robert Schmidt and Gillian Newstead) have been hired and will start next year. We are currently in negotiations with Dr. Newstead to acquire at least 300 cancer cases for our database. We also have an additional 32 normal cases identified. With these images, we should reach our goal of 1000 cases.

It is important that this database be used in such a way so as comparisons between different algorithms can be made. This is the main motivation for creating such a database. To allow for valid comparisons to be made, two things are needed: (i) the exact same cases need to be used to measure performance, and (ii) the same criteria for scoring the results need to be used. To this end, we have divided the database into testing and training cases. Furthermore, we are

preparing recommendations for scoring the results based on some preliminary results in our lab. [1] This information will be released with the database. No other database offers such instructions for its use and thus comparisons of different techniques are still difficult to do. We are in the process of specifying an objective scoring method based on radiologists' input. Once we have the scoring method developed and truth marked, we will begin to distribute the database widely. The study is underway. Preliminary results are inconclusive at this stage.

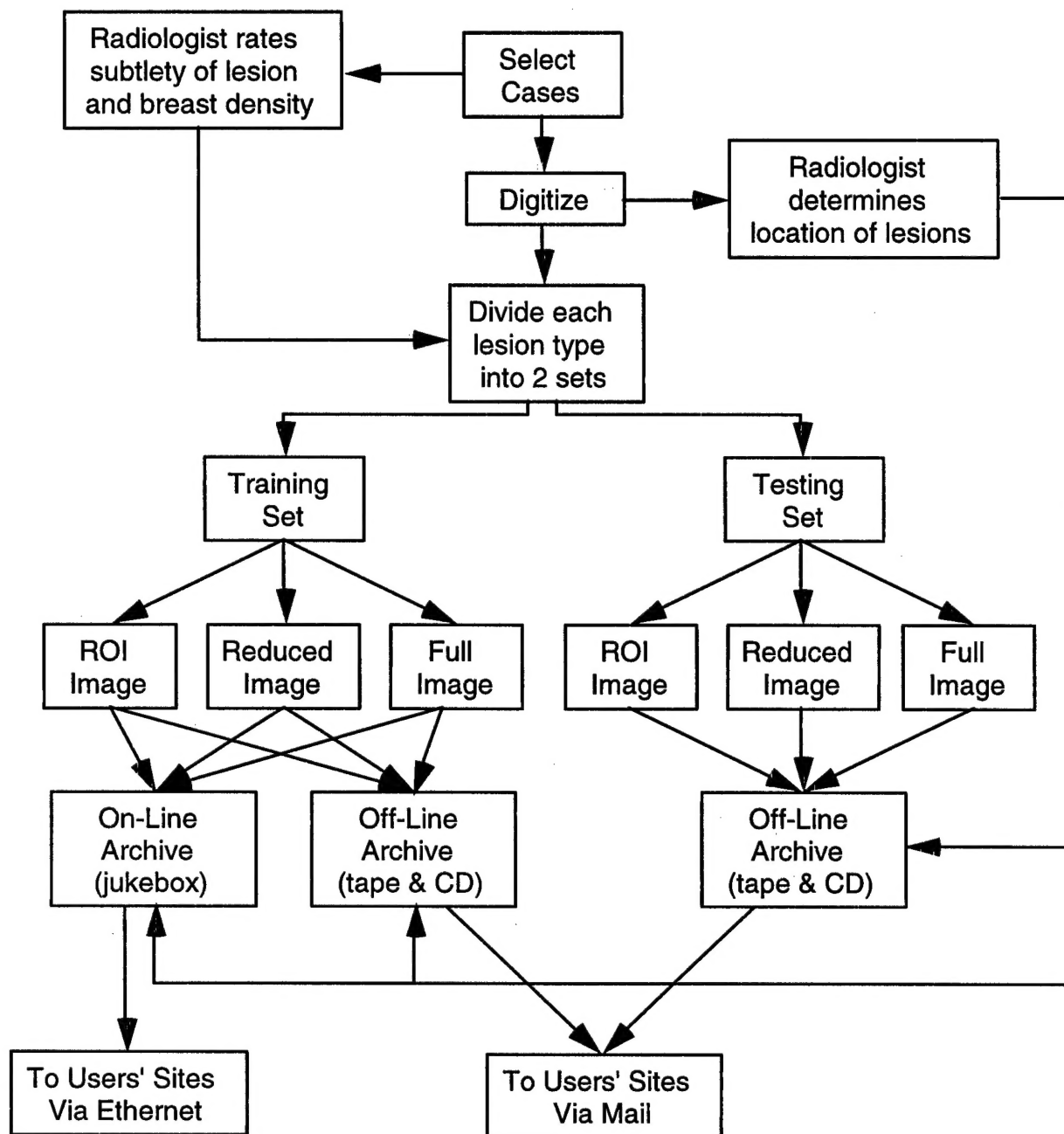


Figure 1. A flowchart of the steps required to collect, digitize, archive, and distribute the mammographic database. The 'Full Image' is the whole digitized mammogram at full resolution. The 'Reduced Image' is a minified version (reduced resolution) of the full image. The 'ROI Image' is a portion of the full image at full resolution.

## Task 2: Establish protocol for transmitting database

We originally considered the ACR/NEMA (DICOM) image format for our database. However, when we began our work, the ACR/NEMA format did not have a module for mammography, and it would have been an extensive project to develop one at that time. Recently a digital mammography module has been approved. We have at this time decided not to use the DICOM format, since many users of our database may not have a DICOM reader.

## Task 3: Maintain database and distribute cases

Maintenance of the database and distribution of the database are at a minimum currently. These tasks will become important shortly as cases go "on-line". Cases are being archived on 4-mm tape and DVD.

## **6. KEY RESEARCH ACCOMPLISHMENTS**

- Collection of 630 mammographic cases

## **7. REPORTABLE OUTCOMES**

### Presentations and Manuscripts:

1. Nishikawa RM, Wolverton DE, Schmidt RE, Johnson RE, Pisano, ED, Hemminger BM: A common database of mammograms for research in digital mammography. U.S. Army Medical Research and Materiel Command Breast Cancer Research Program: An Era of Hope, November, 1997, Washington, DC.
2. Nishikawa RM, Wolverton DE, Schmidt RA, Pisano ED, Hemminger BM, Moody J: A common database of mammograms for research in digital mammography. In: Doi K, Giger ML, Nishikawa RM, and Schmidt RA (eds.), Digital Mammography '96. (Amsterdam: Elsevier Science) 435-438, 1996.
3. Nishikawa RM: Mammographic databases. Breast Disease 10 137-150, 1998.

## **8. CONCLUSIONS**

The development of a common database of mammograms for digital mammography research is underway. We are currently establishing an objective scoring method based on radiologists' input. We will distribute cases and the scoring method in order to insure that meaningful comparisons between different techniques can be made. Such comparisons are currently not possible or are problematic with any existing database.

## **9. REFERENCES**

1. Nishikawa RM, Yarusso LM: Variations in measured performance of CAD schemes due to database composition and scoring protocol. Proc. SPIE 3338: 840-844, 1998.

Table I. Breakdown of cases in the database as of October 1/01.

Type of Lesion	Pathology	# of Cases
Mass	Malignant	116
Mass	Benign	75
Microcalcifications	Malignant	115
Microcalcifications	Benign	87
Asymmetric Density	Malignant	30
Asymmetric Density	Benign	4
Architectural Distortion	Malignant	32
Architectural Distortion	Benign	3
Normal		168
<b>Total</b>		<b>630</b>